DOCUMENT RESUME

ED 439 405 CS 013 922

AUTHOR Hoffman, James V.; Roser, Nancy L.; Salas, Rachel; Patterson, Elizabeth; Pennington, Julie

TITLE Text Leveling and Little Books in First-Grade Reading.

INSTITUTION Center for the Improvement of Early Reading Achievement, Ann Arbor, MI.

SPONS AGENCY Office of Educational Research and Improvement (ED), Washington, DC.
REPORT NO CIERA-R-1-010
PUB DATE 2000-03-31
NOTE 27p.
CONTRACT R305R70004
AVAILABLE FROM CIERA/University of Michigan, 610 E. University Ave., Room 1600 SEB, Ann Arbor, MI 48109-1259. Tel: 734-647-6940. For full text: http://www.ciera.org/ciera/publications/report-series/inquiry-1/1-010.pdf.
PUB TYPE Reports - Research (143)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Beginning Reading; Grade 1; Primary Education; Reading Achievement; Reading Improvement; *Reading Instruction; *Reading Material Selection; *Reading Materials; Reading Research; Validity

ABSTRACT

This paper investigates the validity of two current approaches for estimating text difficulty at the first-grade level--the Scale for Text Accessibility and Support-Grade 1 (STAS-1) and the Fountas/ Pinnell system--and analyzed student performance in relation to each. Subjects, 105 first graders in two schools, used "little books"--easy to read tiny paperbacks that serve as leveled practice materials for beginning readers. The schools are located in an urban area, 90% Latino, low-income, and with 95% of the students qualifying for subsidized school meals. The children were tested and placed into high, middle, and low ability-based reading groups. Each child was assigned a text set to read under either a preview, modeled, or sight reading condition, and their performance was evaluated for accuracy, rate, and fluency. Results indicated that the text leveling procedures these systems employed were largely accurate. Their analysis even suggested potential benchmarks for first-grade performance: 95% accuracy, 80 words per minute, and a fluency level of 3 (on a scale of 1-5) with reading material of mid-first-grade difficulty. More research is needed to determine whether these figures hold for the larger public. Lastly, results indicated that the modeled reading condition seemed particularly supportive of children's subsequent independent reading at any level. (Contains 13 tables of data and 18 references.) (RS)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 439 405

Text Leveling and Little Books in First-Grade Reading

James V. Hoffman, Nancy L. Roser, Rachel Salas,
Elizabeth Patterson, and Julie Pennington
The University of Texas–Austin

BEST COPY AVAILABLE

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

CIERA REPORT #1-010
March 31, 2000

CIERA
Center for the Improvement of Early Reading Achievement

CS 013 922

Text Leveling and Little Books in First-Grade Reading

CIERA REPORT #1-010

James V. Hoffman, Nancy L. Roser, Rachel Salas, Elizabeth Patterson, and Julie Pennington
The University of Texas—Austin

CIERA Inquiry 1: Readers and Texts
What are the characteristics of readers and texts that have the greatest influence on early reading? What are the tasks posed to young readers by currently available beginning reading programs?

In this paper, Hoffman and his colleagues investigated the validity of two current approaches for estimating text difficulty at the first-grade level—the Scale for Text Accessibility and Support-Grade 1 (STAS-1) and the Fountas/Pinnell system—and analyzed student performance in relation to each. They worked with 105 first graders in two schools, using "little books"—easy to read tiny paperbacks that serve as leveled practice materials for beginning readers.

The children were tested and placed into high, middle, and low ability-based reading groups. Each child was assigned a text set to read under either a preview, modeled, or sight reading condition, and their performance was evaluated for accuracy, rate, and fluency.

These readings convinced Hoffman and his colleagues that the text leveling procedures these systems employed were largely accurate. Their analysis even suggested potential benchmarks for first-grade performance: 95% accuracy, 80 words per minute, and a fluency level of 3 (on a scale of 1-5) with reading material of mid-first-grade difficulty. (More research is needed, they stress, to determine whether these figures hold for the larger public.) Lastly, they found that the modeled reading condition seemed particularly supportive of children's subsequent independent reading at any level.

Center for the Improvement of Early Reading Achievement
University of Michigan – Ann Arbor

March 31, 2000

CIERA
Center for the Improvement of Early Reading Achievement

University of Michigan School of Education

610 E University Ave., Rm 1600 SEB
Ann Arbor, MI 48109-1259

734.647.6940 voice
734.763.1229 fax
ciera@umich.edu

www.ciera.org

©2000 Center for the Improvement of Early Reading Achievement.

This research was supported under the Educational Research and Development Centers Program, PR/Award Number R305R70004, as administered by the Office of Educational Research and Improvement, U.S. Department of Education. However, the comments do not necessarily represent the positions or policies of the National Institute of Student Achievement, Curriculum, and Assessment or the National Institute on Early Childhood Development, or the U.S. Department of Education, and you should not assume endorsement by the Federal Government.

Text Leveling and Little Books in First-Grade Reading

James V. Hoffman, Nancy L. Roser, Rachel Salas,
Elizabeth Patterson, and Julie Pennington
The University of Texas—Austin

Soon after the Greeks borrowed and perfected the alphabet, young boys were taught to read. According to some historians, the challenge for the teacher of the day was that there was nothing for children to read between the alphabet and Homer (Guéraud & Jouguet, 1938, as cited in Harris, 1989). The evolving story of reading instruction has been (at least partly) the story of filling the gap between single words and great works with texts intended to support the developing reader (Smith, 1965/1986). For example, a beginning reader in colonial America was offered the *New England Primer,* a text which provided the basic elements of literacy—letters, syllables, and rhyming couplets—all intended to "prime" the child's later reading of the more difficult scriptures. Later, "spellers" were introduced as yet another bridge to more challenging readers (Venezky, 1987).

By the middle of the nineteenth century, arrays of increasingly difficult readers began to be associated with grade levels. By the mid-twentieth century, students' basal series comprised a collection of leveled texts arranged in graduated levels of difficulty, as verified by readability formulas. Typically, a first grader was offered three "preprimers" to build the recognition vocabulary required by the primer, and a "first reader" to stretch the beginner further. The control over the difficulty level for these texts was achieved through careful selection, introduction, and repetition of words (Smith, 1965/1986).

For the beginning reader, standard instruction through the mid-1980s meant practicing in texts that provided for substantial success and a modicum of challenge. In the late 1980s, calls for more authentic literature and less contrived language for beginning reading instruction led basal publishers to abandon their strict leveling procedures and vocabulary control (Wepner & Feeley, 1986) and provide young readers with reproduced trade literature. This "quality literature," with its naturally occurring rhymes, rhythms, and patterns, replaced the carefully leveled vocabulary-controlled texts. Trade book anthologies became the standard basals of the 1990s. The publisher-assigned levels within these basal programs were perhaps driven more by instructional goals and/or thematic integrity than a clear leveling of the

materials according to one or another standard of difficulty (Hoffman et al., 1994).

Classroom research focusing on this shift toward "authentic" literature in first grade revealed mixed effects (Hoffman, Roser, & Worthy, 1998). Although teachers found the new materials more motivating and engaging for their average and above-average readers, they reported difficulties in meeting the needs of their struggling readers with texts so challenging and variable in difficulty. In an attempt to address the need, both basal publishers and others offered supplementary or alternative texts that provided for smaller steps—more refined or narrow levels of text difficulty. Called "little books," these 8-, 12-, or 16-page paperbound texts were designed to provide for practice by combining control (of vocabulary or spelling patterns) with predictable language patterns—the latter an attempt to ensure interest and to include literary traits.

Precise leveling of these little books has been an elusive exercise for both developers and users (Peterson, 1991). Traditional readability formulas, relying on word frequency and syntactic complexity, have not been able to account for variations within the first grade (Klare, 1984). Neither do traditional readability formulas consider features of text support associated with predictable texts (Rhodes, 1981).

Although procedures exist for judging the appropriateness of text-reader match when children are actually reading (e.g., informal reading inventories, running records), the set of teacher tools available for making *a priori* judgments and planning decisions regarding the challenge level of texts is quite limited. Neither are there clearly developed benchmarks for publishers in standardizing the challenge level of the texts they produce. Finally, there are no existing data to validate the text leveling systems that teachers rely upon to array the plethora of practice materials in beginners' classrooms.

The purpose of this study was to investigate the validity of two relatively recent approaches for estimating text difficulty and scaling at the first-grade level: the Scale for Text Accessibility and Support—Grade 1 (STAS-1; Hoffman et al., 1994, 1997) and the Fountas and Pinnell system (1996, 1999). Both attempt to provide teachers with tools that can be used for meeting the goal of putting appropriately leveled practice materials into beginners' hands.

The Scale for Text Accessibility and Support—Grade 1

The first version of the STAS-1 was developed as a tool for investigating the changes in basal texts that had occurred in the transition from the carefully controlled 1980s versions to the literature-based anthologies of the 1990s (Hoffman et al., 1994). In its earliest permutation, STAS-1 consisted of two separate subscales, representing two separate holistic ratings of text.[1] The first subscale focused on decodability features, and the second focused on predictability. Decodability was conceptualized as a factor operating primarily at the word level and affecting the accessibility of text. The assumption was that irregular words—those that do not conform to common ortho-

graphic and sound pattern relationships—and longer words place demands on the developing reader that can make word identification difficult. Predictability was primarily conceptualized as a between-word factor. Larger units of language structure (e.g., rhyming phrases) and other supports (e.g., pictures, familiar concepts) can also support the reader toward accurate word identification and text processing. These two features (decodability and predictability) are independent of one another, at least conceptually (i.e., it is possible to create text that is low in one factor, but high in the other).

To reflect the degree of decodability and predictability demands, the two subscales were arranged on a 5-point rating system, with the lower numbers indicating greater levels of text support available to the reader, and the higher reflecting an increase in challenge level. As with all holistic scales, the categories were designed to represent ranges rather than precise points.

Rating Scale for Decodability

In judging beginners' text for the degree of decodability, the rater focuses on the words in the text, making word-level judgments about the regularity of spelling and phonetic patterns. To judge the degree of decodability, the rater considers the following characteristics:

1. Highly Decodable Text

The emergent or beginning reader would find mostly high-utility spelling patterns (e.g., CVC) in one-syllable words (e.g., *cat, had, sun*). Other words may be short and high frequency (e.g., *the, was, come*). Some inflectional endings are in evidence (e.g., plurals).

2. Very Decodable Text

Beginners still meet mostly high-utility rimes, but useful vowel and consonant combinations appear (e.g., *that, boat, pitch*). Words that seem less decodable are both short and high frequency. Some simple compound words (e.g., *sunshine*) and contractions (e.g., *can't, I'm, didn't*) may appear. In addition, longer, more irregular words occasionally appear as story "features" (e.g., character names, sound words). Although these high-interest words are infrequent, they are often repeated (e.g., *Carlotta, higglety-pigglety*).

3. Decodable Text

Beginners find access to these texts through regularly spelled one- and two-syllable words. Longer words are also composed of regularly spelled units. However, less common rimes may appear (e.g., *-eigh, -irt/-urt*), and more variantly spelled function words (e.g., *their, through*).

4. Somewhat Decodable Text

Beginning readers require more sophisticated decoding skills to access the text, since there is little obvious attention to spelling regularity or pattern. Although most of the vocabulary is still in the one- to two-syllable range, there is greater frequency of derivational affixes (e.g., *dis-, -able*). Some infrequent words and longer nondecodable words appear.

5. Minimally Decodable Text

Beginners' access to the text may depend upon more well-developed skills, since the text includes a plethora of spelling-sound patterns, including longer and more irregularly spelled words (e.g., *thorough, saucer*). There is a full range of derivational and inflectional affixes.

Rating Scale for Predictability

A rater employing the Predictability subscale focuses on a selected text's format, language, and content. To judge the degree of predictability, the rater considers the following characteristics:

1. Highly Predictable Text

Emergent readers can give a fairly close reading of the text after only a few exposures because of the inclusion of multiple and strong predictable features (e.g., picture support, repetition, rhyming elements, familiar events/concepts).

2. Very Predictable Text

An emergent reader can give a fairly close rendering of parts or many sections of the text after only a few exposures to the story. The text includes many features of predictability, but may differ from highly predictable text in both the number and strength of the predictable features.

3. Predictable Text

Emergent or beginning readers can likely make some predictions about language in parts of the text. The text provides attention to predictable features, but only one or two characteristics of predictability may be evident.

4. Somewhat Predictable Text

An emergent or beginning reader might be cued to identification of particular words or phrases and be able to join in on or read portions of the text after several exposures. Attention to predictability is achieved primarily through word repetition rather than through use of multiple features of predictability. A single word or short phrase within more complex text may be the only repeated features.

5. Minimally Predictable Text

An emergent or beginning reader would find no significant support for word recognition as a function of predictable features. The text itself includes few, if any, readily identifiable predictable characteristics or features.

Anchor passages from first-grade materials for each point on both subscales were identified from the materials studied. Again, the anchor passages represented an example within a range of possible texts rather than a precise point.

When the scales were applied to compare the skills-based basal series of the 1980s with the literature-based 1990s series, the newer texts displayed a dramatic increase in both level of predictability and decoding demands. That is, the literature-based series offered more support to young readers (as judged by the texts' predictable features), but this gain was offset by the increased demands for decoding difficult words (i.e., *accessibility*; Hoffman et al., 1993).

The version of the STAS-1 used in this study involved combining the ratings derived from the two subscales. All texts employed in this investigation were rated on the two scales separately using the same feature lists and anchor texts as in the original study. The resulting scores were combined, however, in the following manner:

STAS-1 = .2 (Decodability Rating + Predictability Rating)

Possible scores using this scale range from the lowest rating of .2 (the "easiest" text) to a rating of 2.0 (the "most difficult" text). The midpoint rating of the scale (1.0) is intended to represent, at least theoretically, the level of text that an average first-grade reader could read with 92–98% accuracy, at a rate of 60 to 80 words per minute, and with good sentence level fluency. That is, a rating level of 1.0 might be considered appropriate for middle first-grade text. With the same criteria applied, a top rating of 2.0 is text that the average first grader should be able to read at the end of first grade or the beginning of second. We stress that these are hypothetical benchmarks designed to guide the scaling of texts by teachers and developers for planning and design purposes.

The Fountas/Pinnell Book Gradient System

A widely used system for leveling little books was developed by Fountas and Pinnell (1996). The Fountas/Pinnell Book Gradient System recommends that teachers work together to level texts by developing a set of benchmarks based, for example, on a published leveled set. Other little books and practice materials can then be judged against these prototypes or anchors. The gradient system has 16 levels that stretch between kindergarten and third grade, with 9 levels for kindergarten and first grade. Books are arrayed along a continuum based on a combination of variables that both support readers' developed strategies and give opportunities for building additional ones. The characteristics used to array books in the Fountas/Pinnell system include length, size and layout of print, vocabulary and concepts, language structure, text structure and genre, predictability and pattern of language, and supportive illustrations (p. 114). Descriptions for each of the 9 kindergarten/first-grade levels from the Fountas and Pinnell system are provided in Table 1.

Fountas and Pinnell (1996) maintain that their system is similar in construction to Reading Recovery levels, but differs in the fineness of gradient in arraying books for beginners (see Peterson, 1991). Because the Reading Recovery program is intended to support struggling beginners, it requires even narrower gaps between levels so that teachers can "recognize, record, and build on the slightest indications of progress" (p. 115). As with any system, users are reminded that the real judgments are made in the balance between systems and individual children's needs.

Methodology

The validity of the two systems (STAS-1 and Fountas/Pinnell) was explored in relation to student performance in leveled texts. Our goal was not to pit the systems against one another, but to examine common features of the systems and their effectiveness in leveling texts.

Table 1: Descriptions for Each of Nine Levels From the Fountas and Pinnell System*

K–1 Levels	Descriptions of Texts
Levels A and B	Books have a simple story line, and a direct correspondence between pictures and text. Children can relate to the topic. Language includes naturally occurring structures. Print appears at the same place on each page, and is regular, clear, and easy to see. Print is clearly separated from pictures. There are clear separations between words so children can point and read. Several frequent words are repeated often. Most books have one to four lines of text per page. Many "caption" books (e.g., labeled pictures) are included in Level A. Level B may have more lines and a slightly broader range of vocabulary.
Level C	Books have simple story lines and reflect familiar topics, but tend to be longer (more words, somewhat longer sentences) than Level B books, even though there may be only two to five lines of text per page. Familiar oral language structures may be repeated, and phrasing may be supported by placement on the page. The story is carried by the text, however, and children must attend closely to print at some points because of variation in patterns. Even so, there is still a direct correspondence between pictures and text.
Level D	Stories are a bit more complex and longer than previous levels, but still reflective of children's experiences. More attention to the print is required, even though illustrations continue to support the reading. Most texts at this level have clear print and obvious spacing. Most frequently, there are two to six lines of print per page. There is a full range of punctuation. Words that were encountered in previous texts may be used many times. Vocabulary may contain inflectional endings.
Level E	Stories are slightly more complex and longer; some concepts may be more subtle and require interpretation. Even when patterns repeat, the patterns vary. There may be three to eight lines of text per page, but text placement varies. Although illustrations support the stories, the illustrations contain several ideas. Words are longer, may have inflectional endings, and may require analysis. A full variety of punctuation is evident.
Level F	Texts are slightly longer than the previous level, and the print is somewhat smaller. There are usually three to eight lines of text per page. Meaning is carried more by the text than the pictures. The syntax is more like written than oral language, but the pattern is mixed. The variety of frequent words expands. There are many opportunities for word analysis. Stories are characterized by more episodes, which follow chronologically. Dialogue has greater variety. Punctuation supports phrasing and meaning.
Levels G and H	Books contain more challenging ideas and vocabulary, with longer sentences. Content may not be within children's experiences. There are typically four to eight lines of text per page. As at Level F, literary language is integrated with more natural language patterns. Stories have more events. Occasionally, episodes repeat. Levels G and H differ but the language and vocabulary becomes more complex and there is less episodic repetition.
Level I	A variety of types of texts may be represented. They are longer, with more sentences per page. Story structure is more complex, with more elaborate episodes and varied themes. Illustrations provide less support, although they extend the texts. Specialized and more unusual vocabulary is included.

*The system is summarized from Fountas & Pinnell, 1996, pp. 117–126.

Setting and Participants

Two schools served as research sites for the study. These schools were selected because of their history of collaboration with the local university as professional development schools. The two schools are located in an urban area approximately ten blocks apart. The Spanish language and Hispanic culture are predominant in both schools' neighborhoods. Student enrollment in these schools is 90% Latino, 5% European American, and 5% African American. The community is low-income, with 95% of the students qualifying for free or reduced lunch.

With the exception of monolingual Spanish-speaking students, all first-grade students enrolled in the two elementary schools were considered eligible for participation in the study. A total of 105 first-grade students participated.

Text Selection and Text Characteristics

The texts selected for study were three sets of little books (the designation assigned to the "easy-to-read" tiny paperbacks produced to serve as leveled practice materials for beginning readers). Both of the schools' bookrooms contained organized collections of these little books. The collections had been leveled in the two schools using both the Fountas/Pinnell and the Reading Recovery leveling systems, as interpreted by the schools' Reading Recovery teachers. The two schools operated independently, however, in the development of their book collections and in their interpretation of leveling procedures. Thus, the total bookroom holdings in each school were similar in terms of numbers of books and titles represented, although the ratings of particular books could and did vary between sites. Both collections were readily available to classroom teachers and support staff.

We scrutinized the two book collections with the goal of identifying titles that met two criteria: They (a) appeared in the collections of both schools, and (b) were classified similarly in both schools in adjudged levels of text difficulty. Both schools used a rating system with seven levels to designate appropriate text for developing first-grade readers. As mentioned, each book was labeled in each school with both a letter level (referred to as its Fountas/Pinnell level), and a numerical level (referred to as its Reading Recovery level; see Table 2).

Table 2: Text Difficulty Levels Assigned by Three Systems

Assigned Text Difficulty Levels for this Study	Adapted Reading Recovery Levels	Fountas/Pinnell Levels
1	3/4	C
2	5/6	D
3	7/8	E
4	9/10	F
5	11/12	G
6	13/14	H
7	15/16	I

Once the set of common titles in each library collection had been identified, we randomly selected three book titles for each level of difficulty (from 1 through 7). The three books for each level were then randomly assigned to create three Text Sets (A, B, and C). Thus, each of the three Text Sets consisted of one book from each of the seven levels of difficulty for a total of 21 titles (see Table 3).

Table 3: Titles of Texts in Each Text Set

Book Levels	Text Set A	Text Set B	Text Set C
1	*A Hug is Warm*	*Come On*	*Danger*
2	*Miss Pool*	*No, No*	*Bread*
3	*Jump in a Sack*	*Mrs. Wishy Washy*	*Go Back to Sleep*
4	*Grandpa Snored*	*Meanies*	*Poor Old Polly*
5	*Greedy Cat*	*Caterpillar Diary*	*Grandpa, Grandpa*
6	*Ratty Tatty*	*Mr. Whisper*	*Mrs. Grindy*
7	*Poor, Sore Paw*	*Nowhere, Nothing*	*Mrs. Muddle*

Text Analysis Measures and Ratings

We ran multiple analyses of each of the selected little books. Most of the measures, such as total number of unique words and type/token ratio, have been used in previous studies examining text difficulty (Hiebert & Raphael, 1998; Klare, 1984). All of the words in all 21 texts (with the exception of the title words) were used to calculate these measures (see Table 4).

Table 4: Assessments of Beginners' Texts

Measure	Explanation
Total Number of Words	All text words, exclusive of the title
Total Number of Unique Words	Total number of different words (including inflections and derivations)
Type/Token Ratio	Incidence of unique words in the total text. Calculated by dividing Measure 2 (total number of unique words) by Measure 1 (total number of words)
Readability Index	Produced through the Right-Writer text analysis system. The lowest (default) score for a text with this index is 1.0 (first-grade level)
Syllables Per Sentence	Average number of syllables in each sentence
Syllables Per Word	Average number of syllables in the words in a text
Average Sentence Length	Average number of words per sentence in a text

We calculated the decodability and predictability of each text using the STAS-1 subscales in the following way: At least two members of the research team rated each of the 21 little books for both decodability and predictability. None of the raters' independent judgments varied by more than +/-1 on either scale. Where differences existed in the ratings (e.g., raters split

between scoring 2 and 3), a midpoint rating was assigned (e.g., 2.5). Finally, we created a composite score by summing the two rating scores (decodability + predictability) for each text and multiplying by .2 to reach a rating for each text. This scale had the potential to range from a low of .4 (easiest/most supported passage) to a high score of 2.0 (hardest/least supported passage).

Design

The independent variable of primary interest was the text difficulty, or text leveling factor. However, two other variables were considered as part of the research design: student word recognition level and reading condition (the instructional procedures used to introduce the reading task).

The word recognition skill levels of the 105 students participating in the study were estimated by administering the Word List section of the Qualitative Reading Inventory (QRI; Leslie & Caldwell, 1990). A total word accuracy score was calculated for each student. Students were then assigned to one of three ability groups (High, Middle, or Low) based on their performance on the word list. Approximately the top third of the scores were designated as high, the middle third designated as midrange, and the bottom third designated as low. The average score on the QRI for the high group was 82.9 (SD = 11.7); for the middle group, 33.0 (SD = 16.1); and for the low group, 10.4 (SD = 3.5).

To approximate the varying ways little books are used with young children in classrooms, we also varied the experimental reading conditions to reflect varying levels of support. The first condition was a "Preview" reading condition (similar to guided reading, but without its detail and implied knowledge of the learner) in which a member of the research team provided an opportunity for the student to preview the story under the guidance of the research team member. The student also received some limited help/instruction with potentially challenging words. In the second condition, labeled "Modeled" reading, the text was read aloud to a student by a member of the research team before the student was asked to read it aloud on his or her own. Each student was invited to follow along in the text as it was read aloud, but no specific attention was given to instructing difficult words. This procedure closely matches the classroom instructional procedure called shared reading, but leaves out many of the important support elements described by Holdaway (1979). In the third condition, labeled "Sight" reading, the students were simply invited to read the text aloud without any direct support (see Table 5). In classrooms, the third condition would be most directly comparable to a cold reading of a text.

Students from each of the three ability groups were assigned proportionally to one of the three experimental conditions. Each stratified group of students was assigned to read texts (the ordering of which had been randomized) in one of the three possible classroom simulated instructional conditions. Thus, each student participating in the study, whatever their level of word-reading skill, read all seven texts in one of the sets (either A, B, or C) under one of the three experimental conditions (Preview, Modeled, or Sight). The design was balanced to permit examination of the relationship between any of these variables and student performance.

Table 5: Description of Instructional Support Procedures

Modified Method	Description
Sight Reading	In the sight reading condition, we stated the title of the book while pointing to each word. We explained to the students that they should try their best, read independently, and keep going if they got stuck. After these quick instructions, the students read the book.
Preview (Guided) Reading	In the preview condition, we prepared and followed a script for each book. We created each script based on story elements Fountas and Pinnell emphasize in their guided reading model. After stating and pointing to the title, we gave a short introductory statement of the story's plot. Next, we invited the students to "take a walk through the pictures" and to talk about what they saw in each illustration. During the book walk, we stopped the students one or two times to introduce a vocabulary word or concept. At the end of the book walk, we read a closing statement about the story. After encouraging students to do their best, we invited them to read the book.
Modeled (Shared) Reading	For the modeled reading condition, we stated the title of the book while pointing to each word, and then read the book aloud to the students, pointing as we read. When we were finished reading, we invited the students to read the book.

Procedures

Outside their regular classrooms, each student met with a member of the research team in three separate sessions. All three sessions were tape-recorded. In Session 1, students read from the first five word lists (preprimer through grade 2) of the QRI. During Session 2, the students read the first three texts (the order of which had been randomized) of their assigned Text Set, following the treatment plan they had been assigned. In all treatment conditions, the students read directly from the little books. To be responsive to student frustration with difficult texts, researchers provided help if students paused longer than five seconds for a word regardless of treatment condition.

During Session 3, which took place on the following day, the students read the remaining four little books under the same condition they experienced in Session 2 (Preview, Modeled, or Sight). Most of the students were able to complete the reading of the passages in two sessions of approximately 25 to 30 minutes each, but some students required an additional session.

Data Analysis

Each student's oral reading performance was monitored (by a running record) and examined in relation to three independent variables: (a) students' entering word-reading skill level (high, middle, or low); (b) the reading conditions (Preview, Modeled, or Sight); and (c) the text difficulty (Levels 1 through 7 based on the combined Fountas/Pinnell and Reading Recovery systems). A 3 x 3 x 7 factorial design was employed.

The dependent variables were three aspects of student performance on these texts: accuracy, rate, and fluency. For total word accuracy, we counted the number of words read accurately in each book. To measure fluency, we used the following 5-point scale for rating student performance for each little book read: A student score was one (1) if the reading was halting, choppy, or word-by-word. A score of two (2) indicated some, but infrequent, attempts to read in phrases. A score of three (3) reflected some sentence-level fluency, but some residual choppy performance. Students were assigned a four (4) if their reading was smooth and occasionally expressive; finally, a score of five (5) was reserved for fluent, expressive, interpretive reading.

Table 6: Means for Text Variables/Ratings on the Seven Levels of Text Sets

Features	Text Level Sets						
	1	2	3	4	5	6	7
Decodability	1.9	1.9	2.7	3.5	3.6	4.0	4.0
Predictability	1.7	1.6	2.5	2.9	3.5	4.2	3.6
Readability*	1.0	1.0	1.7	1.2	2.1	1.0	1.4
STAS-1 Scale†	7.2	7.0	10.3	12.7	14.2	16.3	15.2
Fountas/Pinnell Scale (est.)‡	3.0	4.0	5.0	6.0	7.0	8.0	9.0
Reading Recovery Levels (est.)§	3.5	5.5	7.5	9.5	11.5	13.5	15.5
Sentence Length	5.2	6.4	6.9	6.4	7.4	7.3	7.3
Type/Token Ratio	.34	.37	.31	.43	.39	.32	.36
Syllables Per Word	1.1	1.1	1.3	1.9	1.2	1.2	1.2
Syllables Per Sentence	4.6	6.6	7.9	9.2	8.3	8.3	10.5

*The readability estimates were derived from the application of the Right-Writer text analysis system. No estimates are made below the 1.0 level, the default value.

†The STAS-1 Scale was computerd by adding the Predictability and the Decodability ratings and multiplying by .2.

‡The Fountas/Pinnell Levels are estimates made within schools. The scores represent simple transformations from letters (C through I) to numbers (3 through 9).

§The Reading Recovery Levels are estimates made within schools.

This fluency scale was developed and used in a previous research effort (Hoffman et al., 1998), and was found to be highly reliable and highly correlated with student achievement. All members of the research team were trained to apply the fluency rating system. Each researcher either assigned a fluency rating as the child finished each passage, or immediately after the session when listening to the audiotape of the reading. To ensure reliability, a second rater listened to each taped passage and independently assigned a fluency rating. If there were a discrepancy of only one interval on the scale (for instance, a student's performance receiving a 3 from one rater and a 4 from the second), we averaged the two scores (i.e., 3.5). If there was more than a one-point difference, a third rater listened to the taped reading and reconciled the differences between the original two ratings.

To determine rate, we divided each little book into three sections (beginning, middle, and end). Within each section, we located the page spread that contained the greatest number of words. We then counted the number of words read accurately in each of these three text segments. If each student's accuracy rating met the minimum criterion (between 75–80% of words read

accurately), we used the tape of the child's reading to time each selected section, and computed a words-per-minute rate based on the total number of words read accurately. If the student did not reach the minimum accuracy levels for any of the three selected segments, we did not calculate a rate score.

Table 7: Intercorrelations for Text Variables/Ratings

VARIABLES	1	2	3	4	5	6	7	8	9	10
1. Decodability	1.0									
2. Predictability	.77	1.0								
3. Readability	.38	.31	1.0							
4. STAS-1 Scale	.93	.95	.37	1.0						
5. Fountas/Pinnell	.72	.76	.20	.78	1.0					
6. Reading Recovery	.72	.76	.20	.78	1.0	1.0				
7. Sentence Length	.47	.42	.37	.47	.33	.34	1.0			
8. Type/Token Ratio	.45	.41	.23	.45	.02*	.02*	1.0	1.0		
9. Syllables/Word	.37	.27	.42	.33	.35	.37	.37	-.14	1.0	
10. Syllables/Sentence	.61	.64	.36	.66	.59	.59	.65	.31	.47	1.0

*These are the only two correlations that did not achieve levels of statistical significance ($p < .0001$).

Results

The results are described in three ways. First, we present findings based on the inspection and analysis of the little books. Second, we present student performance in relation to text characteristics, reading condition, and student skill levels. Finally, we discuss student performance in relation to the text leveling procedures used in this study.

Examining Text Characteristics

The data presented in Table 6 combine the values from the three texts at each assigned difficulty level. The distributions for the Fountas/Pinnell and the Reading Recovery levels are forced by the design of the study; thus, the increases across levels of difficulty are expected. Problems are evident within Level 5 and Level 7 when considering the text level measures. We attribute these differences to two texts: *Caterpillar Diary* (Level 5 in Text Set B) was rated as more difficult on the STAS-1 than the Fountas/Pinnell or Reading Recovery levels would suggest. *Nowhere, Nothing* (Level 7, Text Set B) was rated as easier on the STAS-1 than either the Fountas/Pinnell or Reading Recovery assigned level.

Table 7 is a correlation matrix of the various text factors. These data suggest that most of the traditional text factors used to differentiate text difficulty (e.g., type/token ratios, syllables per word, syllables per sentence) do not

reflect the same patterns in leveling these beginner texts as do holistic scales. The correlations between the STAS-1 scale and the Reading Recovery and Fountas/Pinnell scales, however, are quite strong.

Table 8: Intercorrelations for Text Variables/Ratings and Student Performance Measures

	Decodability	Predictability	Readability	STAS-1	Fountas/Pinn.	RR Level	Sent. Length	Type/Token	Syll./Word	Syll./Sent.	Stdt. Accuracy	Stdt. Fluency	Stdt. Rate	QRI
Student Accuracy	.21	.15	.25	.25	.20	.20	.26	.08	.11	.17	1.0			
Student Fluency	-.21	-.09*	-.24	-.24	-.19	-.19	-.16	-.10	-.10	-.18	.80	1.0		
Student Rate	-.40	-.34	-.34	-.40	-.30	-.30	-.27	-.17	-.26	-.32	.57	.64	1.0	
QRI	-.03*	-.03*	-.04*	-.03*	.00*	.00*	-.06*	-.04*	-.03*	-.03*	.64	.73	.37	1.0

*These are the only correlations that did not achieve levels of statistical significance ($p < .0001$).

Table 8 presents the performance data for all students in relation to text factors and the scaling systems. On this intercorrelation matrix, the holistic scales reveal a stronger relationship with student performance characteristics than do the isolated text features.

Relating Reading Condition and Student Performance to Text Characteristics

We used an analysis of variance (ANOVA) to examine the relationship between fluency ratings (by ability level) across passage difficulty levels (see Table 9). We also used ANOVA to examine accuracy levels (by ability level) across passage difficulty levels (see Table 10).

Table 9: Fluency Ratings by Ability Across Text Difficulty Levels

Ability Levels	Passage Levels						
	1	2	3	4	5	6	7
High	3.8	3.9	3.9	3.6	3.5	3.6	3.6
Middle	2.9	2.8	2.7	2.6	2.3	2.2	2.2
Low	2.0	2.1	1.8	1.8	1.3	1.3	1.3

	Degrees of Freedom	Sum of Squares	Mean Square	*F* Value	*P* Value
Rdg. Level	2	524.067	262.003	76.126	.0001
Passage Level	6	44.986	7.498	29.438	.0001

Post Hoc Bonferroni/Dunn
All group differences are statistically significant.

Both the fluency and accuracy analyses showed a statistically significant effect for passage level and ability level on performance. In other words, the

more challenging the passages were, the lower the performance was on both variables. To ground these data in a central reference point, we found that the average rate on the middle level set of passages (Level 4 texts) was 95% (*SD* = .06). The average fluency level for the Level 4 texts was 2.7 (*SD* = 1.1). The analyses of the rate data, however, proved problematic. To attempt to make the rate data meaningful, we set a base-level criterion on word-reading accuracy that the student must achieve (80% or better on all three samples) before we would attempt to calculate that student's rate. Because

Table 10: Accuracy Ratings by Ability Across Text Difficulty Levels

Ability Levels	Passage Levels						
	1	2	3	4	5	6	7
High	.98	.98	.98	.96	.96	.96	.96
Middle	.89	.91	.86	.82	.76	.78	.79
Low	.70	.69	.65	.59	.49	.48	.50

	Degrees of Freedom	Sum of Squares	Mean Square	*F* Value	*P* Value
Rdg. Level	2	18.237	9.119	70.082	.0001
Passage Level	6	1.850	.308	26.907	.0001

Post Hoc Bonferroni/Dunn
All group differences are statistically significant.

many of the low group readers, and even some of the middle group readers, did not achieve this level of accuracy, their rate data were not included. The resulting uneven cell sizes made calculating statistical effects impossible. Our analysis of rate, therefore, was limited to a consideration of the performance of middle and high-skill readers. For both groups, we found a statistically significant effect for passage level on rate ($p = .01$) with an average rate of 125 words per minute on the easiest passages (Level 1 texts), an average rate of 82 words per minute on the middle set of passages (Level 4 texts),

Table 11: Fluency Levels by Treatment Condition

Condition	Passage Levels						
	1	2	3	4	5	6	7
Sight	2.5	2.5	2.3	2.3	2.1	2.2	2.1
Preview	2.8	2.7	2.7	2.4	2.2	2.3	2.3
Modeled	3.6	3.7	3.6	3.4	2.9	2.8	2.7

	Degrees of Freedom	Sum of Squares	Mean Square	*F* Value	*P* Value
Condition	2	125.682	62.841	8.641	.0003
Passage Level	6	44.986	7.498	30.364	.0001

Post Hoc Bonferroni/Dunn
Statistically significant differences for the Modeled condition, but not between Sight and Preview.

and an average rate of 80 words per minute on the more difficult passages (Level 6 texts). The average rate for the Level 4 texts was 79.9 words per minute (*SD* = 33.9).

Table 12: Accuracy Levels by Treatment Condition

Condition	Passage Levels						
	1	2	3	4	5	6	7
Sight	.84	.83	.76	.72	.70	.69	.70
Preview	.84	.86	.85	.77	.72	.75	.76
Modeled	.92	.92	.92	.91	.83	.80	.83

	Degrees of Freedom	Sum of Squares	Mean Square	*F* Value	*P* Value
Condition	2	2.243	1.121	3.950	.0222
Passage Level	6	1.850	.308	24.725	.0001

Post Hoc Bonferroni/Dunn
Statistically significant differences for the Modeled condition, but not between Sight and Preview.

An analysis of variance was also used to examine the effects of the experimental support condition on student performance. The results for the fluency and accuracy data are presented in Tables 11 and 12. The differences for the treatment condition were statistically significant. Post hoc analyses suggested that the differences on fluency and accuracy were associated with the Modeled condition. The differences between the Preview and Sight conditions were not statistically significant, although the means suggest a pattern reflecting more success for the Preview over the Sight condition.

Again, because of missing data, our analysis of the rate data was limited to a consideration of the middle and high skill readers. We found no statistically significant effect for reading condition on rate for the middle and high groups, although rate was consistently higher on the easier passage levels (Levels 1 through 3) for the Modeled reading condition.

Predicting Student Performance With Text Measures

A series of multiple regression analyses were conducted using all of the text factor variables and the rating scales to predict student performance. In all of the models, the QRI score was entered first to remove the effects of entering skill level on performance. The best models for predicting performance in the areas of fluency, accuracy, and rate are presented in Table 13. In all three cases, the STAS-1 score combined with the QRI score was the best predictor of student performance.

Table 13: Best Models for Predicting (With Student Ability Included) Using Multiple Regression Analyses

Fluency	*T* Value	*P* Value
QRI Score	31.443	.001
STAS-1 Score	-10.433	.001
Model		R = .770

Total Word Accuracy	*T* Value	*P* Value
QRI Score	24.23	.001
STAS-1 Score	-9.434	.001
Model		R = .688

Rate	*T* Value	*P* Value
QRI Score	13.128	.001
STAS-1 Score	-13.574	.001
Model		R = .619

Discussion

We tested a practical means for arraying texts in ways that can be applied across the hall and across the country with similar results. Our intent was to add to the available studies and varied opinions about what makes a text difficult for beginners. The findings from the study are encouraging. For the research community, the data offer compelling evidence that the kinds of holistic text leveling procedures represented in the Fountas/Pinnell and Reading Recovery systems, as well as the STAS-1, are validated through student performance. For teachers and others involved in materials selection and development, these approaches to text leveling appear useful.

The study yields findings that go beyond the simple ordering of texts by difficulty. As Clay (1991) noted, teachers fail their students if they attend only to the levels of text. The critical issue is the interaction between child and teacher that accompanies a particular reading. To that end, we reproduced approximations of classroom practices—reflecting teachers who Preview or guide reading, those who Model or share reading aloud before children attempt it on their own, and those who offer children the opportunity to read on their own. Although the methodologies we incorporated in this study were much thinner than those provided by the best teachers, they nevertheless registered effect. The oral sharing of the text (reading aloud in an engaging, well-paced way) seemed to particularly support the child's reading that followed.

The data from the STAS-1 analysis also suggest some useful benchmarks that provide opportunities for future research. Specifically, potential benchmarks for first-grade performance may approximate 95% accuracy, 80 words per

minute, and a fluency level of 3 when the children read texts judged as mid-first-grade difficulty (Levels 3–5). We pose these tentative figures with the caution that they represent data gathered in very particularized conditions and contexts.

Although both leveling systems stood up well under the scrutiny of this study, there appear to be distinct strengths associated with each. The STAS-1 scale offers the advantage of unpacking the features of predictablity from decodability. It may also offer a slight advantage in terms of ease of use. The Fountas/Pinnell system offers the advantage of considering a broader array of text features such as the overall length of the text, the number of words per page, and the size of print. Neither system offers a very careful inspection of word-level features. We suspect that the work of Hiebert (1999) in this area may enrich the set of tools available to inspect the texts used for beginning readers.

But What About . . . ?

Several design decisions and unalterable circumstances of our study may have affected the generalizability of results. For example, we observed and measured students enrolled in the professional development schools in which we work. The children in these low-income neighborhoods are predominantly Hispanic and speak English as their second language. We do not know if the patterns we described will generalize more broadly to first-grade students in other settings. Certainly, other categories of readers must be considered before we can suggest general implications.

Second, we focused closely on word-level issues in our measures, narrowly defining children's reading both in our initial measure and in our judgments of book reading performance. Neither did we consider the physical designs of books as a potential support for beginning readers (Peterson, 1991). To examine only a portion of reading prowess is to ignore such important factors as comprehension, engagement, interest, decoding strategies, and children's instructional and life histories. In limiting our focus, we did not discount the wide-ranging goals of effective reading. Rather, in the interest of expediency, we focused on decoding and fluency. In the case of the QRI word recognition test, for example, we selected a manageable instrument with which many children could experience some success—even those whose literacy was just beginning to emerge. We recognize the need to widen the lens in our determination of reading performance, incorporating more of the potential scaffolds available to teachers. However, even without attending to a full range of text features or knowing our participants' individual backgrounds and needs, we found that the accuracy and fluency that these children demonstrated while reading leveled little books gave us insight into their text processing. We can now use these findings in investigating the broader set of reading issues that concern us. The results of this study are in no sense definitive or as clear-cut as we might have hoped. We concur with Hiebert's (1999) admonition that the debate over which text features are useful for beginners has continued for too long in the absence of

empirical data. This investigation linking text factors with student performance is a step toward investigating these issues with a finer lens.

Notes

1. The original scale and procedures are presented in NRRC Technical Report #6 entitled *So What's New in the New Basals? A Focus on First Grade*. The scale presented here has been rearranged in terms of the direction of difficulty. This change was made to permit combining the two scales. Some minor modifications have also been made in the scaling features based on experiences in the training of coders to high levels of reliability.

References

Clay, M. M. (1991). *Becoming literate: The construction of inner control.* Portsmouth, NH: Heinemann.

Fountas, I. C., & Pinnell, G. S. (1996). *Guided reading: Good first teaching for all children.* Portsmouth, NH: Heinemann.

Fountas, I. C., & Pinnell, G. S. (1999). *Matching books to readers: Using leveled books in guided reading, K-3.* Portsmouth, NH: Heinemann.

Harris, W. V. (1989). *Ancient literacy.* Cambridge, MA: Harvard University Press.

Hiebert, E. H. (1999). Text matters in learning to read. *The Reading Teacher, 52,* 552-566.

Hiebert, E. H., & Raphael, T. E. (1998). *Early literacy instruction.* Fort Worth, TX: Harcourt Brace.

Hoffman, J. V., McCarthey, S. J., Abbott, J., Christian, C., Corman, L., Curry, C., Dressman, M., Elliott, B., Matherne, D., & Stahle, D. (1993). *So what's new in the new basals? A focus on first grade* (National Reading Research Report No. 6). Athens, GA, and College Park, MD: National Reading Research Center.

Hoffman, J. V., McCarthey, S. J., Abbott, J., Christian, C., Corman, L., Dressman, M., Elliot, B., Matherne, D., & Stahle, D. (1994). So what's new in the "new" basals? *Journal of Reading Behavior, 26,* 47-73.

Hoffman, J. V., McCarthey, S. J., Elliott, B., Bayles, D. L., Price, D. P., Ferree, A., & Abbott, J. A. (1997). The literature-based basals in first-grade classrooms: Savior, Satan, or same-old, same old? *Reading Research Quarterly, 33* (2), 168-197.

Hoffman, J. V., Roser, N. L., & Worthy, J. (1998). Challenging the assessment context for literacy instruction in first grade: A collaborative study. In C. Harrison & T. Salinger (Eds.), *Assessing reading: Theory and practice* (pp 166-181). London: Routledge.

Holdaway, D. (1979). *The foundations of literacy.* Exeter, NH: Heinemann.

Klare, G. (1984). Readability. In P. D. Pearson, R. Barr, M. Kamil, & P. Mosenthal (Eds.), *Handbook of reading research* (pp. 681-744). New York: Longman.

Leslie, L., & Caldwell, J. (1990). *Qualitative reading inventory.* New York: HarperCollins.

Peterson, B. (1991). Selecting books for beginning readers: Children's literature suitable for young readers. In D. E. DeFord, C. A. Lyons, & G. S. Pinnell (Eds.), *Bridges to literacy: Learning from Reading Recovery* (pp. 115-147). Portsmouth, NH: Heinemann.

Rhodes, L. K. (1981). I can read! Predictable books as resources for reading and writing instruction. *The Reading Teacher, 36,* 511-518.

Smith, N. B. (1965/1986). *American reading instruction.* Newark, DE: International Reading Association.

Venezky, R. L. (1987). A history of the American reading textbook. *The Elementary School Journal, 87,* 247–281.

Wepner, S., & Feeley, J. (1986). *Moving forward with literature: Basals, books, and beyond.* New York: Maxwell Macmillan International.

About CIERA

The Center for the Improvement of Early Reading Achievement (CIERA) is the national center for research on early reading and represents a consortium of educators in five universities (University of Michigan, University of Virginia, and Michigan State University with University of Southern California and University of Minnesota), teacher educators, teachers, publishers of texts, tests, and technology, professional organizations, and schools and school districts across the United States. CIERA is supported under the Educational Research and Development Centers Program, PR/Award Number R305R70004, as administered by the Office of Educational Research and Improvement, U.S. Department of Education.

Mission. CIERA's mission is to improve the reading achievement of America's children by generating and disseminating theoretical, empirical, and practical solutions to persistent problems in the learning and teaching of beginning reading.

CIERA Research Model

The model that underlies CIERA's efforts acknowledges many influences on children's reading acquisition. The multiple influences on children's early reading acquisition can be represented in three successive layers, each yielding an area of inquiry of the CIERA scope of work. These three areas of inquiry each present a set of persistent problems in the learning and teaching of beginning reading:

CIERA INQUIRY 1
Readers and Texts

Characteristics of readers and texts and their relationship to early reading achievement. What are the characteristics of readers and texts that have the greatest influence on early success in reading? How can children's existing knowledge and classroom environments enhance the factors that make for success?

CIERA INQUIRY 2
Home and School

Home and school effects on early reading achievment. How do the contexts of homes, communities, classrooms, and schools support high levels of reading achievement among primary-level children? How can these contexts be enhanced to ensure high levels of reading achievement for all children?

CIERA INQUIRY 3
Policy and Profession

Policy and professional effects on early reading achievement. How can new teachers be initiated into the profession and experienced teachers be provided with the knowledge and dispositions to teach young children to read well? How do policies at all levels support or detract from providing all children with access to high levels of reading instruction?

www.ciera.org

CIERA

Center for the Improvement of
Early Reading Achievement

CIERA is a collaboration of
University of Michigan
University of Virginia
Michigan State University
with
University of Minnesota
University of Southern California

University of Michigan School of Education
610 E University Av, Rm 1600 SEB
Ann Arbor, MI 48109-1259
734.647.6940 voice
734.763.1229 fax
ciera@umich.edu

www.ciera.org